Broken scaling in the Forest Fire Model

Gunnar Pruessner and Henrik Jeldtoft Jensen
Department of Mathematics, Imperial College, 180 Queen's Gate, London SW7 2BZ, UK
gunnar.pruessner@physics.org and h.jensen@ic.ac.uk
(Dated: February 1, 2008)

We investigate the scaling behavior of the cluster size distribution in the Drossel-Schwabl Forest Fire model (DS-FFM) by means of large scale numerical simulations, partly on (massively) parallel machines. It turns out that simple scaling is clearly violated, as already pointed out by Grassberger [P. Grassberger, J. Phys. A: Math. Gen. **26**, 2081 (1993)], but largely ignored in the literature. Most surprisingly the statistics not seems to be described by a universal scaling function, and the scale of the physically relevant region seems to be a constant. Our results strongly suggest that the DS-FFM is not critical in the sense of being free of characteristic scales.

1. INTRODUCTION

The Drossel-Schwabl Forest Fire Model [1] (DS-FFM) is one of the paradigms of non-conservative SOC [2]. Its importance comes primarily from the fact that the model has non-conservative microdynamics. It therefore answered the question whether conservation is necessary for criticality in driven systems.

The claim that the DS-FFM is critical comes from the fact that it shows powerlaw-like behavior for several geometrical properties of the dissipation events. The average size of these is divergent in the so-called SOC limit, where all timescales get separated so that the rate of the external drive becomes infinitely slow and the total inflow diverges. If one assumes stationarity this is trivial to prove [1, 3, 4]. However, as usual in numerical simulations, it is not possible to investigate the model in the limit of divergent drive ($\theta \to 0$ in the notation below), as finite size limits the correlation length and therefore destroys any possible criticality [4]. It is remarkable that most of the literature available for this model is mainly concerned with finding critical exponents and identifying supposedly critical quantities. It seems that no authors question whether the model is critical at all and if so in which sense. In this paper we carefully investigate the "scaling function" of the cluster size distribution and show that it is indeed an open question whether the model is truly critical: Not only is there no way to prove its criticality, there is also numerical evidence that the model does not become scale free.

2. DEFINITION OF THE MODEL AND METHODS

The model has been described several times and in great detail elsewhere [1, 3, 4]. Therefore, the description presented here is rather succinct. The model is defined on a d-dimensional lattice of linear size L, where each lattice site can be in one of two states, "occupied" or "empty". The lattice is then updated according to the following rules:

- Driving: Choose randomly $(1/\theta)$ sites, one after the other. If its state is "empty" turn it into "occupied".
- Relaxation: Choose one site at random. If it is empty continue with the first step. Otherwise make "empty" all the sites in the cluster the site chosen belongs to. In this case the update is considered to be successful. Continue with the first step.

Here a cluster is defined in the usual fashion as the set of occupied sites which are connected via nearest neighbor interactions, i.e. two sites belong to the same cluster if they are nearest neighbors or there is a path between them along sites, which belong to the same cluster. We have applied periodic boundaries in all our simulations and restrict ourselves to the two dimensional square lattice.

The cluster removed in the second step is called the "burnt cluster". To measure the overall distribution of clusters within the system, one usually measures the size of the burnt cluster [4], the distribution of which is biased by a factor s. To see that, we define $n(s)$ to be the ensemble-averaged, site-normalized density of clusters of size s in the system. Then, the probability that a randomly chosen site is connected to a cluster of size s is $sn(s)$, as in standard percolation [5]. This distribution is probably the most important in the model. Other quantities are the distribution of the burning time, which is defined as the maximum Manhattan distance (shortest path on the square lattice) from the initially chosen site of a burnt cluster to all other sites in the same cluster, and the correlation functions as defined and discussed in [6]. In this paper we solely concentrate on the distribution $n(s)$.

Using a new implementation of the model [7] we are able to simulate the system on very large scales and at the same time keep track of the *entire* distribution $n(s)$, instead of measuring the biased distribution $sn(s)$, as done usually [4]. Between two updates the changes in $n(s)$ are only of the order $(1/\theta)$, so it is a highly correlated quantity. However, by using standard cluster labeling techniques [8], it is possible to calculate the full histogram $n(s)$ essentially without increasing the computing time, which depends almost exclusively on θ and is essentially independent of the system size. Compared to the stan-

dard simulation, we gain up to two orders of magnitude in performance [19]. A similar method was recently introduced for standard percolation [9]. Using the same amount of computing time the results are significantly less noisy than those of the standard implementation (for example [10], which we have used as reference to check the validity of our results). Large system sizes enable us to rule out any finite size effects as described below. The results have been partly cross-checked using a different random number generator (all results presented here make use of `ran2` from [11], and for checking `ran1` from [11] has been used).

Finite size effects have been ruled out by the following direct method: For each value of $(1/\theta)$ a significantly larger system was simulated with exactly the same value for $(1/\theta)$. The linear size L was typically increased by a factor 2. The smallest systems we used were $L = 1000$, the largest $L = 32000$. By comparing the histograms of different system sizes in conjunction with the standard deviation calculated for them, it is possible to decide whether a system size is affected by finite size effects or not. Compared to other simulations published [12, 13], which also claim not to suffer from finite size effects, our system sizes are huge. This simple method of "redoing" all simulations and using lattices which are much larger than actually needed has the obvious disadvantage of being inefficient, but there is probably no more a direct way of identifying finite size effects [14]. This waste of computing power is overcompensated by the efficiency of the algorithm and self-averaging [15].

3. RESULTS

The focus of this paper is the presumably universal scaling function of the distribution $n(s)$. Similar to the correlation function one expects

$$n(s;\theta) = s^{-\tau}\mathcal{G}(s/s_0(\theta)) \tag{1}$$

if "simple scaling" applies, which is already known not to be the case in the presence of finite size effects [13]. The L-dependence of this quantity is omitted in the following wherever the context allows it. It is worthwhile to note that this is usually the *definition* of the exponent τ. The function $\mathcal{G}$ is the (presumably) universal scaling function, which depends only on the ratio $s/s_0(\theta)$, where $s_0(\theta)$ is the *only* scale of the distribution. This scale depends only on the external parameters, in our case θ. Simple scaling allows another scale, namely the lower cutoff, but this is fixed or at least bounded.

The function $\mathcal{G}(x)$ is usually smooth for small values of x, therefore it does not make a big difference whether we investigate $n(s)$ or

$$\begin{aligned}\widetilde{n}(s;\theta) &= \frac{n(s;\theta)}{n(1;\theta)} \\ &= \frac{\mathcal{G}(s/s_0(\theta))}{\mathcal{G}(1/s_0(\theta))}s^{-\tau} \quad .\end{aligned} \tag{2}$$

That this particular choice of the normalization does not affect the overall results can be seen in Tab. I, where the absolute value of $n(1,\theta)$ is listed for different values of θ. Also shown in this table is the first moment of the distribution or the average density of trees, which is defined as

$$\rho = \sum_s sn(s;\theta) \tag{3}$$

and the second moment of the distribution,

$$\langle s\rangle = \frac{\sum_s s^2 n(s;\theta)}{\sum_s sn(s;\theta)} \quad , \tag{4}$$

which is the average size of the cluster connected to a randomly chosen occupied site.

Before presenting the actual results, we first discuss the numerical quality of the results.

3.1. Avoiding finite size effects

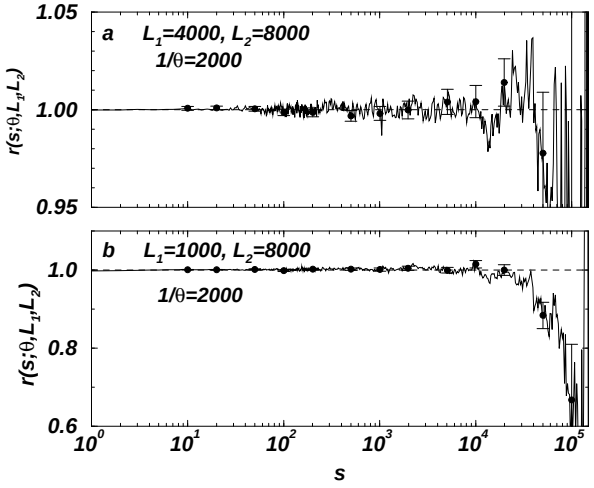

FIG. 1: Ratio $r(s;\theta,L_1,L_2) = \widetilde{n}(s;\theta,L_1)/\widetilde{n}(s;\theta,L_2)$ with $(1/\theta) = 2000$ for two pairs L_1, L_2 with error-bars (one standard deviation; the error-bars as well as the data shown are exponentially binned). The data are from short runs (10^6 updates for statistics). Finite size effects have been considered negligible under the condition that (almost all) error-bars for this ratio have covered 1 (marked by a dashed line) in the relevant range. a) $L_1 = 4000$ and $L_2 = 8000$: Almost no finite size effects, the deviation from 1 is probably due to noise. Note the fine scale of the ordinate. b) $L_1 = 1000$ and $L_2 = 8000$: Systematic, strong finite size effects for $s \gtrapprox 10^4$. The scale of the ordinate is five times larger than in a). Data of this quality have been dismissed.

Throughout this paper we initially performed $5 \cdot 10^6$ successful updates (as defined in Sec. 2) as transient (and therefore rejected them) and the same number for producing statistics, apart from runs for calculating error-bars, where only 10^6 updates has been used for statistics, see below. It is known that the transient can be

TABLE I: Static quantities for different choices of L and $(1/\theta)$. The estimation of the standard deviation of the tree density ρ, $\sigma^2(\rho) = \langle\rho^2\rangle - \langle\rho\rangle^2$, where the average runs over the ensemble, is unfortunately based only on a small subset of the configurations produced, and in the case of the large systems ($L \geq 16000$) only on a fraction of the lattice. However, it is apparent that it behaves like $1/L$, as expected for a system without finite size effects. The density of clusters of size 1, $n(1)$, serves as the normalization of $\widetilde{n}$. The average cluster size is denoted by $\langle s\rangle$, for definition see (4), but due to a truncation in the histogram for some of the simulations in the range $2000 \leq (1/\theta) \leq 16000$, the number presented is actually the average size of the burnt cluster. In the stationary state it is - apart from statistical fluctuations - also given by $(1-\langle\rho\rangle)/(\theta\langle\rho\rangle)$ [4]. Values of $(1/\theta)$ and L printed in bold indicate results shown in Fig. 3, the other results are only for comparison. All data are based on $5\cdot10^6$ (successful) updates (s. Sec.2) for the transient and statistics, apart from those printed in italics which are based on short runs ($5\cdot10^6$ updates for the transient and $1\cdot10^6$ updates for statistics).

$(1/\theta)$	L	$\langle\rho\rangle$	$\sqrt{\sigma^2(\rho)}$	$n(1)$	$\langle s\rangle$	$\frac{1-\langle\rho\rangle}{\theta\langle\rho\rangle}$
125	*1000*	0.3797	0.0060	0.04553	204.07	204.18
125	1000	0.3798	0.0058	0.04552	203.81	204.15
125	*4000*	0.3798	0.0014	0.04553	203.88	204.10
125	**4000**	0.3798	0.0015	0.04552	203.77	204.10
250	*1000*	0.3876	0.0083	0.04451	395.03	395.06
250	1000	0.3875	0.0082	0.04452	394.08	395.15
250	*4000*	0.3877	0.0022	0.04454	394.97	394.89
250	**4000**	0.3877	0.0021	0.04454	395.29	394.91
500	*1000*	0.3932	0.0117	0.04380	764.73	771.75
500	1000	0.3932	0.0119	0.04380	764.81	771.77
500	*4000*	0.3934	0.0031	0.04382	771.12	770.88
500	**4000**	0.3934	0.0030	0.04382	771.90	770.87
1000	*1000*	0.3972	0.0169	0.04328	1495.36	1517.91
1000	1000	0.3971	0.0168	0.04328	1490.05	1518.00
1000	*4000*	0.3976	0.0043	0.04331	1510.85	1515.00
1000	**4000**	0.3976	0.0043	0.04331	1513.13	1514.81
1000	*8000*	0.3976	0.0021	0.04332	1510.10	1514.91
2000	*4000*	0.4005	0.0060	0.04296	2976.34	2993.35
2000	**4000**	0.4005	0.0062	0.04297	2990.50	2993.15
2000	*8000*	0.4006	0.0030	0.04297	2995.67	2992.56
4000	*4000*	0.4026	0.0089	0.04273	5929.24	5935.91
4000	4000	0.4025	0.0089	0.04273	5930.97	5938.03
4000	*8000*	0.4026	0.0048	0.04274	5931.32	5935.15
4000	**8000**	0.4026	0.0046	0.04273	5935.36	5936.47
8000	*4000*	0.4040	0.0135	0.04255	11786.97	11799.72
8000	4000	0.4041	0.0135	0.04255	11788.90	11799.07
8000	*8000*	0.4041	0.0068	0.04257	11801.31	11795.98
8000	**8000**	0.4041	0.0068	0.04257	11792.82	11795.38
16000	*4000*	0.4052	0.0199	0.04244	23430.01	23481.82
16000	*8000*	0.4054	0.0096	0.04243	23466.93	23467.22
16000	8000	0.4054	0.0098	0.04243	23446.10	23465.64
16000	**16000**	0.4054	0.0052	0.04245	23449.31	23466.57
32000	16000	0.4066	0.0075	0.04232	46443.83	46701.82
32000	**32000**	0.4066	0.0032	0.04233	46731.44	46698.51
64000	**32000**	0.4078	0.0042	0.04220	91148.64	92952.40

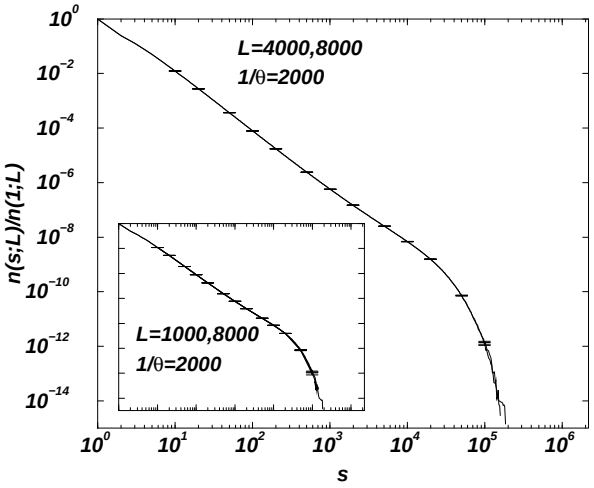

FIG. 2: The binned histogram $\widetilde{n}(s;\theta,L)$ for two different values of L and fixed θ as in Fig. 1a. In this plot the two histograms are virtually indistinguishable. However, note that the deviations shown on Fig. 1b would also hardly be visible in this type of plot, as shown in the inset.

very long [6] (note that the time unit in [6] is expressed in our units by multiplying it with $(1/\theta)/(\rho L^2)$), but in all cases presented the number of initial steps seemed to be more than sufficient. Numerical checks indicate that the cluster size distribution is very stable against the size of the transient, i.e. even a transient, which is presumably too short, still produces reasonable results for $n(s)$.

All systems have been initialized by a random independent distribution of trees with density 0.41.

The standard deviation of the binned histogram is not completely trivial to calculate. In particular, its computation requires a significant amount of CPU time, and was therefore only calculated for the smaller system sizes (up to $L = 8000$) and in shorter runs (only 10^6 updates for statistics, but $5\cdot10^6$ for transient). We resorted to visual examination for the larger systems when comparing $\widetilde{n}(s;\theta,L)$ for different system sizes. Fig. 1a and b show the ratio of $\widetilde{n}(s;\theta,L)$ for two different system sizes. A deviation of this ratio from 1 indicates a difference in the statistics and therefore the presence of finite size effects. Fig. 1a shows a typical case we accepted as reasonable agreement. Here $L_1 = 4000$ and $L_2 = 8000$ do not seem to differ for $(1/\theta) = 2000$. Fig. 1b shows a case of finite size corrections we have dismissed (note the different scales in the two graphs). It differs from Fig. 1a only by $L_1 = 1000$.

Fig. 2 illustrates the strong agreement of $\widetilde{n}(s;\theta)$ at the same value of θ for the same two different sizes L as in Fig. 1a. The two sets of data are virtually indistinguishable, but in this kind of plot it is also almost impossible to see a difference between the data of $L_1 = 1000$ and $L_2 = 8000$, as shown in the inset of Fig. 2. This is also the case with the rescaled data below, and the use of very large systems throughout this paper might there-

fore be "overcautious" in avoiding finite size effects, although such large sizes are obviously required for an accurate *quantitative* analysis of this model. However, when it comes only to qualitative analysis, such a judgment seems to be justified. On the other hand, an increase in system size hardly increases the computing time and affects "only" the memory requirements, which forced us to implement the algorithm for parallel machines. The side effect of using multiple CPUs at the same time is a significant reduction of the simulation time especially for large values of $(1/\theta)$, a fact which compensates the complications of parallel coding.

Another indicator for the absence of finite size effects is the scaling of the standard deviation of ρ: If the lattice can be split into independent parts, i.e. if subsets of the lattice can be considered as independent, the standard deviation of ρ should scale like $1/L$ for different values of L at given $(1/\theta)$. Such a behavior can be seen in Tab. I, although the standard deviation of ρ could be calculated only roughly. This might explain the slight mismatch for $(1/\theta) = 32000$, $L = 16000, 32000$.

For the highest values of $(1/\theta)$ we could not yet do the comparison to another system, so the curve for the largest value of $(1/\theta)$ in Fig. 3 is dotted, as their quality is not known. However, it is reasonable to assume that it is not affected by finite size scaling.

3.2. The scaling function

Comparing the different histograms $\widetilde{n}(s;\theta)$ for different values of $(1/\theta)$ in a plot enables us not only to find the exponent τ, but also to find the universal function $\mathcal{G}$ as defined in Eqn. 1. A rough, naive estimate of τ is given by $\widetilde{n}(s;\theta)$ fitted against $s^{-\tau}$, which gives a value of $\tau^* \approx 2.1$ in our case. Plotting now $\widetilde{n}(s;\theta)s^{\tau^*}$ double logarithmically should allow us to find the "true" value of τ by performing a data collapse, i.e. choosing τ^* in such a way that horizontal shifts (corresponding to the choice of the scale $s_0(\theta)$ in the scaling function) make all curves collapse. This is shown in Fig. 3, where $\tau^* = 2.1$ was chosen so that the maxima for the second bumps are almost equally high: denoting their position on the abscissa for each value of θ by $s_{\max}(\theta)$, we have chosen τ^* such that

$$\widetilde{n}(s_{\max}(\theta);\theta)\; s_{\max}^{\tau^*}(\theta) \approx \text{const.} \quad . \tag{5}$$

According to (1) the constant is simply the maximum value of $\mathcal{G}$, namely $\mathcal{G}(s_{\max}(\theta)/s_0(\theta))$, where the value of the argument is therefore the same for all θ.

The value of τ^* is close to (but not within the error of) the exponent found in the literature, $\tau = 2.14(3)$ [3, 4] ($\tau = 2.15(2)$ in [16], $\tau = 2.159(6)$ in [6]), which is shown in the same figure for comparison. However, it is impossible to force the minima (see the down pointing marks in Fig. 3) to the same height while maintaining the constraint that the maxima remain aligned, i.e. these

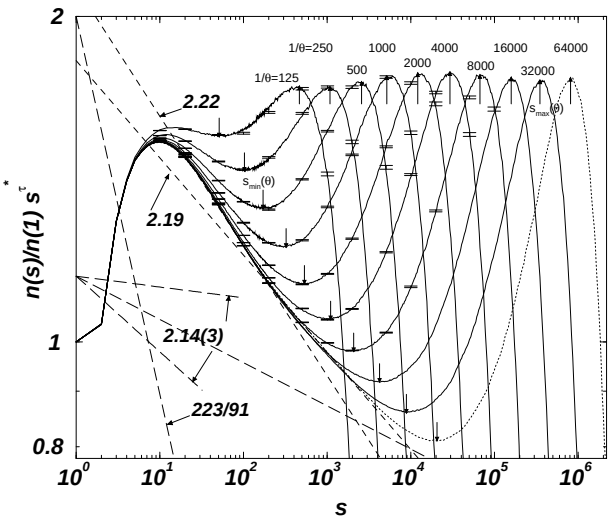

FIG. 3: The rescaled and binned histogram $\widetilde{n}(s;\theta)s^{\tau^*}$, where $\tau^* = 2.10$ for $(1/\theta) = 125, 250, 500, \cdots, 32000, 64000$ (as indicated) in a double logarithmic plot. The linear size L is chosen according to the bold printed entries in Tab. I and large enough to ensure absence of finite size effects. The error-bars are estimated from shorter runs. The rightmost histogram (in all figures dotted, $(1/\theta) = 64000$) could not be crosschecked by another run, see text. Maxima are marked by arrows pointing upwards, minima are marked by arrows pointing downwards. The dashed lines belong to different exponents, whose value is specified as the sum of the slope in the diagram and τ^*, i.e. a horizontal line would correspond to an exponent 2.1. The shortly dashed line are estimated exponents for different regions of the histogram (2.22 within approx. $[20, 200]$ and 2.19 within $[200, 2000]$), the other exponents are from literature, namely 2.14(3) in [3, 4] and $223/91 \approx 2.45$ in [12]. Since it was impossible to relate these exponents to any property of the data, the exact position of the lines associated with them was chosen arbitrarily.

minima cannot be a feature of the same universal scaling function. Otherwise (1) would hold and the quantity

$$\widetilde{n}(s_{\min}(\theta);\theta)\; s_{\min}^{\tau^*}(\theta) \quad , \tag{6}$$

where $s_{\min}(\theta)$ denote the position of the minima, would assume the same value for all θ, because they are local minima of $\mathcal{G}$, which is supposed to be the same for all θ.

Since these minima cannot be included in the simple scaling defined in (1), they must be explicitly excluded by introducing a lower cutoff, so that simple scaling supposedly sets in only above these cutoffs, excluding especially the minima. However, such a lower cutoff would apparently have to diverge for $(1/\theta) \to \infty$ – something that is certainly beyond any established concept of scaling. Even when accepting this peculiar scaling behavior, a data collapse for the second bump still seems to be unsatisfactory, as shown in Fig. 4.

If one accepts a divergent lower cutoff of the scaling function, one has to face the fact that this would describe

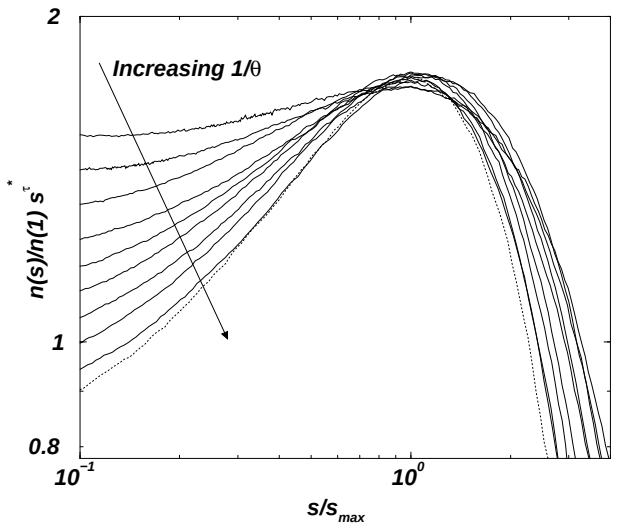

FIG. 4: The rescaled and binned histogram $\widetilde{n}(s;\theta)s^{\tau^*}$, versus $s/s_{\max}(\theta)$, where $\tau^* = 2.10$ for $(1/\theta) = 125, 250, 500, \cdots, 32000, 64000$ in a double logarithmic plot. The scales $s_{\max}(\theta)$ by which the histograms have been shifted are the maxima marked in Fig. 3, so that a data collapse would be possible. The arrow indicates the order of the data in increasing $(1/\theta)$.

the behavior of $\widetilde{n}$ in a region, which becomes physically less and less interesting in the limit $(1/\theta) \to \infty$, because the vast majority of events are situated at small s and as the second bump moves out to infinity, the scaling function hence covers a smaller and smaller part of $\widetilde{n}$. However, only a region of $\widetilde{n}$ which covers a non-vanishing fraction of events can be *physically relevant*.

Concentrating now on the behavior of $\widetilde{n}$ up to the minimum (see arrows pointing downwards in Fig. 3), one finds that this region is also badly described by a function like (1). First of all, the question of which region is supposedly described by the function needs to be answered. A unique lower cutoff and a θ dependent upper cutoff needs to be found. At first view it looks appealing to choose these two marks such that they cover the set of data, where the curves fall on top of each other. In this case the lower cutoff would be 1 and the upper cutoff, s_{naive}, would have a value smaller than the minima marked by downwards pointing arrows in Fig. 3. However, this would be described by a function like

$$\widetilde{n}(s;\theta) = f(s)\mathcal{G}(s/s_{\text{naive}}) \tag{7}$$

rather than (1). Note the *parameter independent* function $f(s)$ describing the shape of the curve, while $\mathcal{G}(s/s_{\text{naive}})$ is a sharp cutoff function. Eqn. 7 does not allow for an exponent, $f(s)$ is an arbitrary function. Writing it as

$$f(s) = s^{-\tau}(a_0 + \text{higher order corrections}) \tag{8}$$

defines τ to be the steepest descent of this part of the curve and gives a value between $\tau_{\text{stp.}} = 2.22$ and $\tau_{\text{stp.}} = 2.19$ (see Fig. 3).

This concept appears to be rather naive – on the other hand, it is hard to assume that (1) can still hold: it would correspond to (7) with $f(s)$ replaced by s^{τ}, which is a straight line in a double logarithmic plot. Therefore (1) can apply only to a region in Fig. 3 where the data that fall on top of each other form a straight line. Those features not already collapsing would then collapse when properly tilted (choosing the right τ) and shifted (choosing the right s_0). Introducing a lower cutoff at $s = 10$ and discarding the data for $(1/\theta) \leq 2000$ then leads to a data collapse in a narrow range as shown in Fig. 5. It is worthwhile mentioning that even for some $10 < s < 200$, namely for values of s between the squares and the filled circles, none of the data collapse. The exponent used in this "collapse" is $\tau_{\text{stp.}} = 2.19$, as mentioned above.

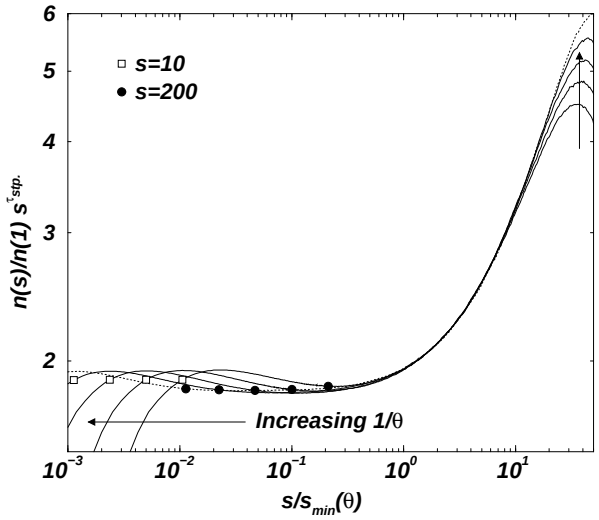

FIG. 5: The rescaled and binned histogram $\widetilde{n}(s;\theta)s^{\tau_{\text{stp.}}}$, versus $s/s_{\min}(\theta)$, where $\tau_{\text{stp.}} = 2.19$ for $(1/\theta) = 4000, 8000, 16000, 32000, 64000$ in a double logarithmic plot. The scales $s_{\min}(\theta)$ by which the histograms have been shifted are slightly different from the minima marked in Fig. 3, to make the collapse as good as possible. The squares and the filled circles mark $s = 10$ and $s = 200$, respectively, for orientation and relation to other figures. The arrows indicate the order of the data in increasing $(1/\theta)$.

By considering the function $f(s)$ it becomes apparent that $\widetilde{n}$, and therefore the model, cannot be scale free: it depends on the fixed, microscopic scale $s = 1$. This entails that it is always possible to tell $(1/\theta)$ by looking only at the *shape* of $\widetilde{n}$; a diagram showing only this shape, without any scales on the axes, reveals $(1/\theta)$, since a scale is intrinsically given by the features of $f(s)$. One would only need to rescale and tilt it until it fits the plot Fig. 3 and one could identify $(1/\theta)$. Only if $f(s)$ were scale free, i.e. a straight line in a double logarithmic plot, would this not be possible.

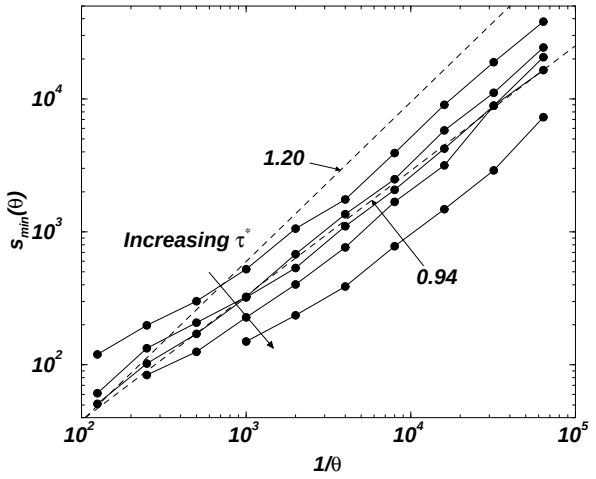

FIG. 6: The position of the minimum in the binned and rescaled histogram for different values of $\tau^* = 2.04, 2.08, 2.10, 2.12, 2.16$. The exponents shown in the plot are for comparison only.

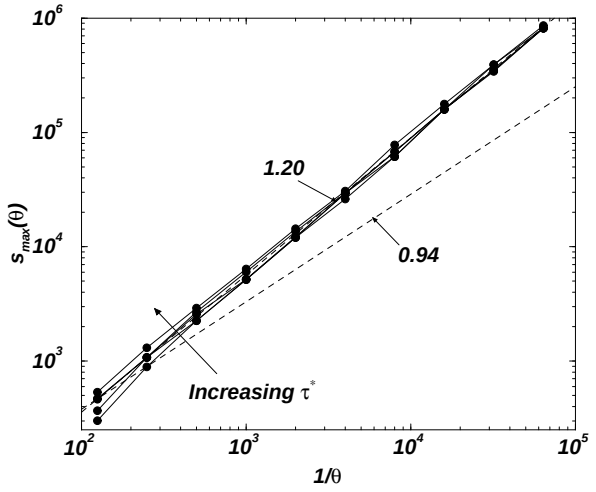

FIG. 7: The position of the maximum in the binned and rescaled histogram for different values of $\tau^* = 2.04, 2.08, 2.10, 2.12, 2.16$. The exponents shown in the plot are for orientation only.

3.3. Two length scales

That $\widetilde{n}$ contains features to define at least two scales, which apparently diverge in $(1/\theta)$ with different exponents, becomes clear when analyzing the scaling of the minima and maxima as marked in Fig. 3, using the definitions

$$s_{\min}(\theta) \propto (1/\theta)^{x_{\min}} \quad (9)$$

$$s_{\max}(\theta) \propto (1/\theta)^{x_{\max}} \quad . \quad (10)$$

Of course, the exact position of the extrema of $\widetilde{n}(s;\theta)s^{\tau^*}$ depends on its tilt, i.e. on the choice of τ^*. However, their *scaling* in $(1/\theta)$ does not depend strongly on this choice. In particular $x_{\min}$ and $x_{\max}$ are different for all choices of τ^*. A plot of $s_{\min}(\theta)$ versus $(1/\theta)$ for different values of τ^* is shown in Fig. 6. For small values of $(1/\theta)$ the minimum is not pronounced enough to survive for large values of τ^*, so these curves do not give a data point. Using a linear fit of $\log s_{\min}(\theta)$ versus $\log(1/\theta)$ of the minimum as found in the rescaled (τ^*) and binned histogram, gives an "exponent" between $x_{\min} = 0.93$ and $x_{\min} = 0.98$. The same procedure applied to the maxima gives an "exponent" in the range $x_{\max} = 1.18$ and $x_{\max} = 1.22$, shown in Fig. 7. One may expect that $x_{\min}$ tends towards $x_{\max}$ for decreasing τ^*, as $s_{\min}$ increases and might enter the scaling region of $s_{\max}$, but neither "exponent" exhibits a systematic variation, and the quality of the fit certainly suffers from the rough procedure that searches for the extrema in the *binned* histogram. This is unfortunately necessary because of statistical fluctuations, in conjunction with the absence of error-bars for all data points.

The scale of the clusters, $s_{\min/\max}$ is related to the correlation length ξ by the fractal dimension μ, i.e. (see [4])

$$s_{\min/\max} \propto \xi^{\mu_{\min/\max}} \quad . \quad (11)$$

Since $\xi \propto (1/\theta)^{\nu}$, one should expect $\nu = x_{\min/\max}/\mu_{\min/\max}$. The minima are supposed to be dominated by smaller, fractal events (see [12]), so $\mu_{\min} = 1.96(1)$ [4] and therefore $\nu_{\min} \in [0.47, 0.50]$. The maxima are more likely to be dominated by compact fires, so $\nu_{\max} \in [0.59, 0.61]$. It is unclear how the two exponents $\nu_{\min/\max}$ are related exactly to the exponents of the two correlation lengths found by Honecker and Peschel [6] for the connected correlation function $\nu = 0.576(3)$ and for the tree-tree correlation function $\nu = 0.541(4)$.

4. DISCUSSION

Prima facie the DS-FFM looks like a percolation process, and one might naively think that it is indeed a percolation process which organizes itself to the critical density: sites are occupied randomly and independently and (at least in the thermodynamic limit) there is only one cluster which is removed with non-vanishing probability, namely the largest. In this way the density of occupied sites is automatically reduced below the percolation threshold whenever the threshold is reached. It is puzzling how remarkably close the tree density in the DS-FFM is to the density of *empty* sites in critical site percolation on a square lattice ($\rho_{\mathrm{FFM}} \approx 0.4078$ and $1-\rho_{\mathrm{perc}} = 0.40725379(13)$ [9] respectively). However, the removal process involved in the DS-FFM introduces spatial correlations which are not present in standard percolation. These correlations are expressed, for example, in the form of a patchy tree density distribution [12].

The purpose of this paper is *not* to add yet another model to the enormous zoo of SOC models. However, in

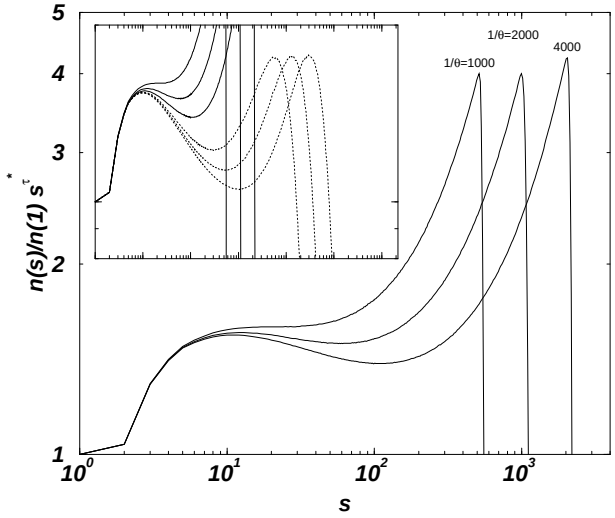

FIG. 8: The rescaled and binned histogram $\widetilde{n}(s;\theta)s^{\tau^*}$ (again $\tau^* = 2.10$), for a modified model, where the largest cluster in the system is removed after each driving step, for $(1/\theta) = 1000, 2000, 4000$ (as indicated) with linear sizes $L = 2000, 2000, 4000$. The inset shows the same data on the scale of Fig. 3 for comparison. The data for $(1/\theta) = 1000, 2000, 4000$ of the original model as shown in Fig. 3 are dotted. The peculiar behavior of the different height-scaling of the minimum and the maximum is again visible.

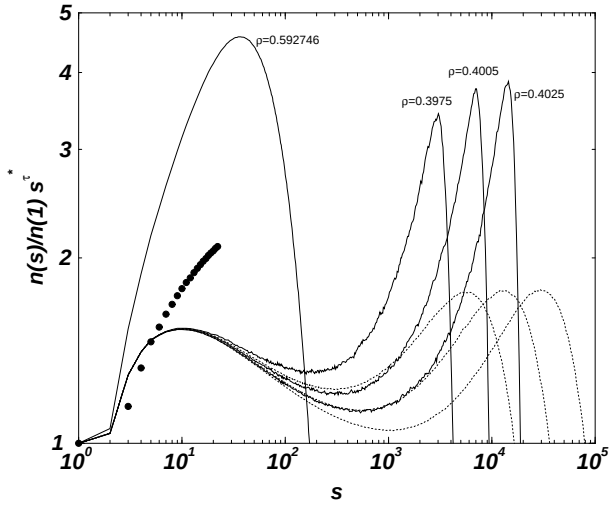

FIG. 9: The rescaled and binned histogram $\widetilde{n}(s;\rho)s^{\tau^*}$ (again $\tau^* = 2.10$), for a modified model, where the largest cluster in the system is removed in each relaxation step and the corresponding number of trees is filled back into the system afterwards. The three small values of ρ chosen, $\rho = 0.3975, 0.4005, 0.4025$ correspond (up to the last digit) to the values of the tree density for $(1/\theta) = 1000, 2000, 4000$ respectively, see Tab. I. The linear size was $L = 1000, 2000, 4000$. The corresponding data of the original model are shown dotted. The peculiar behavior of the different height-scaling of the minimum and the maximum is again visible (a correct tilt τ^* would make it even more pronounced), but disappears obviously for $\rho = \rho_{\rm perc}$ – for these data it is relevant to mention that $\widetilde{n}(s)$ was measured *after* the relaxation. The filled circles show the exact results for the lattice animals [5, 17, 18] at $\rho = \rho_{\rm perc}$.

order to investigate certain features of the given model and identify underlying mechanisms, it makes sense to modify it slightly. The outcome for the histogram of the DS-FFM modified such that the *largest* cluster is removed after each driving step is shown for a few values of $(1/\theta)$ in Fig. 8. The distinctive feature of a minimum which scales differently from the maximum is again present, as the peaks of the maxima have approximately the same height, while the height of the local minima varies among different values of θ. The inset of this figure shows the histogram on the same scale as Fig. 3 together with the data of the original model (dotted) with the corresponding values of θ. One can understand that they do not fall on top of each other, because the relaxation rule in the modified model erases much larger clusters than in the original model.

Fig. 9 shows a second modification of the model, where again the *largest* cluster is removed during relaxation and in addition the driving is changed such that the density of trees, ρ, is the same before each relaxation; the trees removed during the relaxation are just filled in randomly afterwards. This model differs from standard percolation only by its updating scheme [20]. In order to compare the outcome with the original model, the values of ρ have been chosen close to the values given in Tab. I. Indeed, the feature of different scaling of the extrema is still present, but it disappears completely if the density is increased to $\rho_{\rm perc} = 0.592746$ [9], which is shown in the same figure as the large bump. This curve does not vary much if a much smaller system size is simulated at this density, so we expect it essentially to be free of finite size corrections. Since it represents a *correlated* percolation process, it is just consistent that this bump does not cover the exact results for the lattice animals of standard percolation [5, 17, 18] at $\rho = \rho_{\rm perc}$ shown as filled circles in Fig. 9. The dotted graphs in the figure show the corresponding data of the original model. Again they do not match apart from the region of very small s. Unfortunately the simulations of the so-modified model are very expensive in CPU time, because the mass of the largest cluster needs to be refilled after each relaxation, so that only 50.000 updates for transient and statistics could be done.

Since the feature of different scaling survives the modifications described above, it seems reasonable to assume that any relaxation rule that favors the largest cluster leads to the peculiar behavior. Its disappearance at high densities can be explained by the extremely small cutoff in the distribution, which leads to a domination of the statistics by very small clusters, while a single, enormously large one dominates the burning (the average size of the burnt cluster for $\rho = \rho_{\rm perc}$ was 355811). However,

much more careful and detailed investigations of models like the modification described above are required to gain a full understanding of the underlying mechanisms. In particular, this should include a modification of the rules such that the feature disappears.

Honecker and Peschel [6] have calculated the correlation length not only for the probability that two sites belong to the same cluster, but also for the probability that two sites are occupied at all. The correlation function for the latter is of course a δ peak in ordinary percolation, as there are no spatial correlations for the distribution of occupied sites by construction. However, in the DS-FFM the correlation length for this quantity, ξ, is finite and seems to diverges when approaching the critical point. It is highly remarkable that Honecker and Peschel conclude from their simulations that this correlation length diverges slightly *slower* than the correlation length of the probability for two sites to belong to the same cluster, ξ_s. This seems to indicate that for sufficiently large scales the spatial correlation of the occupation probability becomes arbitrarily small, so that on sufficiently large scales the DS-FFM occupation is uncorrelated and therefore tends to standard percolation. In other words, it seems to be possible to rescale or "renormalize" the DS-FFM to make the occupation correlation arbitrarily small. This would introduce higher order interactions, as known from standard real space renormalization group and would explain the difference in critical density between the rescaled DS-FFM and standard percolation. However, if this "mapping" is valid, one should find the exponent for the divergence of ξ_s/ξ to be the same as in standard percolation, but this is precluded by numerics.

It has been suggested at least twice [6, 12], that the DS-FFM is a superposition of cluster distributions $n_{\mathrm{perc}(s,p)}$ of standard percolation for a whole range of concentrations p, weighted by a certain distribution function $w(p)$, i.e. $\int_0^1 dp\, w(p) n(s,p)$. Obviously such an assumption neglects spatial correlations. We recall the following result from standard percolation theory [5],

$$n(s,p) \propto s^{-\tau} \mathcal{C}(-s/(p-p_c)^{-1/\sigma}) \quad , \tag{12}$$

where $\mathcal{C}$ denotes the cutoff function and the exponents σ and τ_{perc} have their standard definitions. Assuming that the weighting function $w(p)$ is analytic around the critical concentration in standard percolation, p_c, (12) leads to

$$\int_0^1 dp w(p) n(s,p) \propto s^{-(\tau_{\mathrm{perc}}+\sigma)} \quad . \tag{13}$$

This gives rise to an exponent $\tau = 223/91 \approx 2.45$, however, this is definitely not supported by numerics (see Fig. 3).

It remains completely unclear how to characterize the scaling of the DS-FFM in two dimensions. Apparently it is not a mere superposition of two simple scalings, as recently speculated [12]. Moreover the model does not seem to be scale free as described above and it does not seem to be possible to identify a unique power law behavior of the cluster size distribution. Nevertheless *effective* power law behavior over restricted regions has clearly been produced by the model, making it potentially relevant to observation.

All we can conclude is that *the DS-FFM is not critical* in the sense of simple scaling. It reminds us that a divergent moment (here $\langle s \rangle$, the second moment) can be regarded as a unique sign of emergent scale invariance only if we are certain that one single scale is sufficient to characterize the system. If there is more than one relevant scale, different properties of the system might depend on different scales which may or may not diverge.

Acknowledgments

The large scale simulations represent the backbone of this paper. Andy Thomas at the Department of Mathematics at Imperial College was eager to support us technically, making machines available and upgrading the systems in our department such that parallel computing became a pleasure. A great deal of the results in this paper was possible only because of his work. We thank him very much.

This work partly relies on resources provided by the Imperial College Parallel Computing Centre. We want to especially thank K. M. Sephton for his support.

Another part of this work was possible only because of the generous donation made by "I-D Media AG, Application Servers & Distributed Applications Architectures, Berlin". We especially thank M. Kaulke and O. Kilian for their support.

G.P. wishes to thank A. Honecker, I. Peschel and K. Schenk for very helpful communication, as well as N. R. Moloney for providing the lattice animal data and for proofreading.

The authors gratefully acknowledge the support of EPSRC.

[1] B. Drossel and F. Schwabl, Phys. Rev. Lett. **69**, 1629 (1992).
[2] H. J. Jensen, *Self-Organized Criticality* (Cambridge University Press, New York, NY, 1998).
[3] S. Clar, B. Drossel, and F. Schwabl, J. Phys. C: Condens. Matter **8**, 6803 (1996).
[4] S. Clar, B. Drossel, and F. Schwabl, Phys. Rev. E **50**, 1009 (1994).
[5] D. Stauffer and A. Aharony, *Introduction to Percolation Theory* (Taylor & Francis, London, 1994).

[6] A. Honecker and I. Peschel, Physica A **239**, 509 (1997).
[7] G. Pruessner and H. J. Jensen, in preparation.
[8] J. Hoshen and R. Kopelman, Phys. Rev. B **14**, 3438 (1976).
[9] M. E. J. Newman and R. M. Ziff, Phys. Rev. Lett. **85**, 4104 (2000).
[10] A. Honecker, *Program for simulating the two-dimensional forest-fire model*, http://www-public.tu-bs.de:8080/~honecker/software/forest2d.html (1997).
[11] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C* (Cambridge University Press, New York, NY, 1992), 2nd ed.
[12] K. Schenk, B. Drossel, and F. Schwabl (2001), preprint cond-mat/0105121.
[13] K. Schenk, B. Drossel, S. Clar, and F. Schwabl, Eur. Phys. J. B **15**, 177 (2000).
[14] G. Pruessner, D. Loison, and K.-D. Schotte, Phys. Rev.B **64**, 134414 (2001).
[15] Ferrenberg, Landau, and Binder, J. Stat. Phys. **63**, 867 (1991).
[16] P. Grassberger, J. Phys. A: Math. Gen. **26**, 2081 (1993).
[17] M. F. Sykes and M. Glen, J. Phys. A: Gen. Phys. **9**, 87 (1976).
[18] S. Mertens, J. Stat. Phys. **58**, 1095 (1990).
[19] As the standard deviation (actually the estimator of the standard deviation of the estimated mean; this quantity includes the correlation time) decreases with the square root of the computing time, one has to compare the products of the square root of the computing time and the standard deviation. In the algorithm presented here, the ratio of the computing time varies between 1.4 and 2.1
[20] Actually, it also differs from standard percolation because it fixes the number of occupied sites rather than simply the probability of being occupied. However, this difference becomes irrelevant for sufficiently large systems.